

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

JC971 U.S. PRO
09/904147
07/11/01


別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

2000年 7月 11日

出願番号
Application Number:

特願 2000-210599

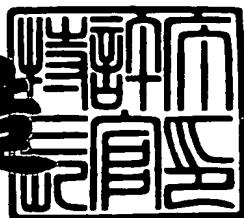
出願人
Applicant(s):

インターナショナル・ビジネス・マシーンズ・コーポレーション

2000年12月22日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特 2000-3107298

【書類名】 特許願
【整理番号】 JP9000133
【提出日】 平成12年 7月11日
【あて先】 特許庁長官 殿
【国際特許分類】 G10L 3/00
【発明者】
【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビ
ー・エム株式会社 東京基礎研究所内
【氏名】 森 信介
【発明者】
【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビ
ー・エム株式会社 東京基礎研究所内
【氏名】 西村 雅史
【発明者】
【住所又は居所】 神奈川県大和市下鶴間1623番地14 日本アイ・ビ
ー・エム株式会社 東京基礎研究所内
【氏名】 伊東 伸泰
【特許出願人】
【識別番号】 390009531
【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレ
ーション
【代理人】
【識別番号】 100086243
【弁理士】
【氏名又は名称】 坂口 博
【復代理人】
【識別番号】 100104880
【弁理士】
【氏名又は名称】 古部 次郎

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【選任した代理人】

【識別番号】 100106699

【弁理士】

【氏名又は名称】 渡部 弘道

【選任した復代理人】

【識別番号】 100100077

【弁理士】

【氏名又は名称】 大場 充

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0004480

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 単語予測方法、音声認識方法、単語予測装置、音声認識装置、コンピュータ・システム、記憶媒体およびプログラム伝送装置

【特許請求の範囲】

【請求項1】 予測すべき単語を含む複数の単語からなる文の文構造を特定し、

特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係にある単語および／または単語列を選択し、

選択された前記単語および／または単語列に基づいて前記予測すべき単語を予測することを特徴とする単語予測方法。

【請求項2】 前記単語および／または単語列は、文構造において部分解析木を構成するものであることを特徴とする請求項1に記載の単語予測方法。

【請求項3】 選択された前記単語および／または単語列が複数ある場合に、複数の前記単語および／または単語列に基づいて単語の予測を行なうことを特徴とする請求項1に記載の単語予測方法。

【請求項4】 予測すべき単語と他の単語との係り受けの関係を特定するステップと、

特定された前記他の単語から前記予測すべき単語を予測するステップと、を備えたことを特徴とする単語予測方法。

【請求項5】 前記係り受けの関係は係り受けの方向を含み、前記予測すべき単語がその前に存在する単語に係る場合があることを特徴とする請求項4に記載の単語予測方法。

【請求項6】 予測すべき単語と他の単語および／または単語列との係り受けの関係が複数通り存在し、その各々について単語を予測することを特徴とする請求項4に記載の単語予測方法。

【請求項7】 文頭から j 番目の単語までを含む文の構造を特定するステップと ($j = 0, 1, 2, \dots$)、

前記 j 番目の単語までを含む文の構造から $j + 1$ 番目の単語に係る単数または複数の部分解析木を特定するステップと、

前記単数または複数の部分解析木に基づき前記 j + 1 番目の単語を予測するステップと、

予測された j + 1 番目までを含む文の構造とその確率値を求めるステップと、

以上の各ステップを文末の単語まで実行した後に、前記確率値が最大となる文構造および単語列を音声認識の結果として選択するステップと、

を備えたことを特徴とする音声認識方法。

【請求項 8】 学習されたテキストデータが記述された辞書を記憶した記憶手段と、

予測すべき単語を含む複数の単語からなる文の文構造に基づいて前記予測すべき単語と係り受けの関係にある単語および／または単語列を選択し、かつ選択された前記単語および／または単語列に基づいて前記予測すべき単語を予測する予測手段と、

を備えたことを特徴とする単語予測装置。

【請求項 9】 学習されたテキストデータが記述された辞書を記憶した記憶手段と、

予測すべき単語と係り受けの関係を有する単語および／または単語列を選択し、選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する予測手段と、

を備えたことを特徴とする単語予測装置。

【請求項 10】 前記辞書は、学習されたテキストデータに基づく所定の単語または単語列に対して他の所定の単語が出現する頻度を記憶していることを特徴とする請求項 9 に記載の単語予測装置。

【請求項 11】 前記予測手段は、予測された単語を含む単語列の確率値を前記頻度に基づき計算することを特徴とする請求項 10 に記載の単語予測装置。

【請求項 12】 前記予測手段は、前記確率値が最大となる単語列を予測結果として選択することを特徴とする請求項 11 に記載の単語予測装置。

【請求項 13】 アナログ音声入力信号をデジタル信号に変換処理する音響処理手段と、

音の特徴を学習した音響モデルを記憶した記憶手段と、

所定の単語または単語列に対する他の所定の単語の出現頻度を記述した辞書を記憶した記憶手段と、

前記デジタル信号について前記音響モデルおよび前記辞書を用いて確率値を計算して最も確率の高い単語を入力した音声として認識する認識手段と、を備え

前記認識手段は、予測すべき単語を当該単語を含む文の構造に基づき予測するとともに、前記出現頻度を用い予測された単語を含む文について前記確率値を計算することを特徴とする音声認識装置。

【請求項14】 予測すべき単語と他の単語および／または単語列との係り受けの関係を特定し、前記係り受けの関係を有していると特定された単語および／または単語列を用いて前記予測すべき単語を予測する手段と、

前記手段により予測された単語および／または単語列、ならびに係り受けの関係を表示する表示手段と、

を備えたことを特徴とするコンピュータ・システム。

【請求項15】 コンピュータ・プログラムが記憶されかつコンピュータで読み取り可能な記憶媒体であって、

前記コンピュータ・プログラムは、

予測すべき単語を含む複数の単語からなる文の文構造を特定し、

特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係にある単語および／または単語列を選択し、

選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する処理をコンピュータに実行させるものであることを特徴とする記憶媒体。

【請求項16】 コンピュータ・プログラムが記憶されかつコンピュータで読み取り可能な記憶媒体であって、

前記コンピュータ・プログラムは、

予測すべき単語と他の単語および／または単語列との係り受けの関係を特定するステップと、

特定された前記他の単語および／または単語列から前記予測すべき単語を予測するステップと、をコンピュータに実行させるものであることを特徴とする記憶

媒体。

【請求項17】 コンピュータに、予測すべき単語を含む複数の単語からなる文の文構造を特定し、特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係にある単語および／または単語列を選択し、選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する処理を実行するためのプログラムを記憶する記憶手段と、

前記記憶手段からプログラムを読み出して前記プログラムを送信する送信手段と、

を備えたことを特徴とするプログラム伝送装置。

【請求項18】 コンピュータに、予測すべき単語と他の単語および／または単語列との係り受けの関係を特定するステップと、特定された前記他の単語および／または単語列から前記予測すべき単語を予測するステップとを実行するためのプログラムを記憶する記憶手段と、

前記記憶手段からプログラムを読み出して前記プログラムを送信する送信手段と、

を備えたことを特徴とするプログラム伝送装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識装置およびその方法に関するものであり、特に、人の自然な発話を認識して文章化し、テキストデータを作成する音声認識装置およびその方法に関するものである。

【0002】

【従来の技術】

音響モデルと言語モデルとを用いて音声認識を行なう統計的な手法は知られている。例えば、「A Maximum Likelihood Approach to Continuous Speech Recognition(L.R. Bahl他, IEEE Trans. Vol. PAMI-5, No.2, 1983, March)」や、「単語を認識単位とした日本語の大語彙連続音認識（西村他、情報処理学会論文誌、第40巻、第4号、1999年4月）」に紹介されている。

その概略について説明すると、生成された文章としての単語列Wが発話され、それが音響処理部において音響処理されて得られた信号からその特徴量Xが抽出され、そのXおよびWを用いて、以下の式にしたがって、最適と考えられる認識結果W'が出力され、文章が構成される。つまり、単語列Wが発声されたときの当該特徴量(X)の出現確率P(X|W)とW自身の出現確率(P(W))の積が最大(arg max)となる単語列が認識結果W'として選択される。

【0003】

【数1】

$$W' = \underset{w}{\operatorname{argmax}} P(w | X) \underset{w}{\operatorname{argmax}} P(w) P(X | w)$$



【0004】

ここで、前者の確率P(X|W)を求めるために音響モデルが用いられ、その確率の高い単語が認識の候補として選択される。一方、後者の確率(P(W))を近似するためによく用いられるものが言語モデルである。

従来の言語モデルでは、通常、直近の単語列を履歴として利用している。その1例として、N-gramモデルがある。これは、N個の連続した単語組の出現確率から文全体、すなわち単語列Wの出現確率を近似する方法であり、次式のように定式化されている。

【0005】

【数2】

$$\begin{aligned} P(w) &= P(w_0)P(w_1|w_0)P(w_2|w_0w_1) \times \dots, P(w_n|w_0w_1, \dots, w_{n-1}) \\ &= P(w_0)P(w_1 | w_0) \prod_{i=2}^n P(w_i | w_{i-2}w_{i-1}) \end{aligned}$$

【0006】

以上の式では、次の単語w[n]の出現確率が直前のN-1個の単語にのみ影響を受けると仮定する。Nの値は種々のものが考えられるが、その有効性と必要とする学習データのバランスからN=3がよく用いられ、本式もN=3の場合を記

述している。N = 3 の場合、tri-gramあるいは3-gramと呼ばれている。以下、n 個の単語からなる単語列Wのn 番目の単語をw [n] のように表現することになると、ここでは当該N - 1 個（つまり2 個）という条件の元での単語w [n] の出現確率、つまり $P(w[n] | w[n-2] w[n-1])$ として単語列Wの出現確率が計算される。ここで、この式において、| の左(w [n])は予測（または認識）の対象となる単語を示し、| の右(w [n-2] w [n-1])はその条件となる2つ前、1つ前の単語を示す。さまざまな単語w [n] についてそれぞれの出現確率 $P(w[n] | w[n-2] w[n-1])$ は予め用意されたテキストデータより学習しておき、辞書としてデータベース化して保存しておく。例えば、文の先頭に「単語」という語が出現する確率は0.0021、その後に「検索」が続く確率は0.001、というようにして保存される。

【0007】

tri-gramモデルを簡単な文章を例にして説明する。

文章例は「さらに小選挙区の」であり、次の「導入」を予測する例である。図8 (a) に予測前を、また図8 (b) に予測後の状態を示す。図8 (a) に示すように、この文章は「さらに」、「小」、「選挙」、「区」および「の」の5つの単語から構成され、予測する単語は「?」で示している。また、図8中の矢印は、各単語間の係り受けの関係を示している。前述の様に、tri-gramでは常に直前の2つの単語から次の単語を予測する。したがって、この例では図8 (a)において実線で囲む「区」と「の」とから「導入」を予測する。

しかし、文構造によっては、直前の2つの単語から次の単語を予測するtri-gramが最良の手法でないことがある。例えば、図9に示す、「何がいま政治の再生のために」という文章の次の単語を予測する場合である。tri-gramによれば、図9 (a) に示すように「ため」と「に」とから「必要」を予測するが、これら以外にも「何が」や「いま」など構造的に関係のある単語を考慮することで予測精度が向上することが予想される。

【0008】

【発明が解決しようとする課題】

直前の2つの部分解析木の主辞から次単語の予測を行なうモデルをChelba & J

elinekが提案している。Chelba & Jelinek のモデルでは、単語は出現の順に予測される。したがって、 i 番目の単語を予測する段階では $i-1$ 番目までの単語とその構造が確定している。この状態から、まず、直前の2つの部分解析木の主辞から次単語と品詞を順に予測する。このとき、直前の2つの部分解析木の主辞と予測される単語との係り受けの関係は考慮しない。単語を予測した後に当該単語を含めた文構造を更新する。したがって、直前の2つの単語から次の単語を予測する tri-gram に比べて予測精度の向上が期待できる。ところが、彼らのモデルでは、単語間の係り受けの関係の有無に関わらず、直前の2つの部分解析木の主辞から単語を予測する。そのために、文構造によっては逆に予測精度を下げる可能性がある。tri-gram で示した「さらに小選挙区の」という文章に基づいて説明する。

図10に示すように、「さらに小選挙区の」という文章は、「さらに」と「小選挙区の」という2つの部分解析木から構成され、各々の主辞は図10 (a) で実線で囲む「さらに」および「の」である。したがって、Chelba & Jelinekの方法によれば、図10 (b) に示すように直前の2つの主辞である「さらに」および「の」から、次の単語「導入」を予測する。そして、「導入」を予測後、図10 (c) に示すように「導入」を含めた文の構造を予測する。この構造の予測は、図中の矢印で示しているように単語間の係り受けを含む。ここで、「さらに」は、「導入」と係り受けの関係ないので、「導入」という単語の予測の役に立たないばかりか、予測精度を下げることになる可能性があることに注意しなければならない。

また、「何がいま政治の再生のために」という文章の例では、以下のようになる。この文章は、「何が」、「いま」および「政治のために」という3つの部分解析木から構成され、各々の主辞は、「が」、「いま」および「に」である。図11 (a) で実線で囲むように、直前の2つの主辞は「いま」と「に」である。したがって、図11 (b) に示すように、「必要」は「いま」と「に」から予測される。「必要」の予測後に、図11 (c) に示すように、「必要」を含めた文章の構造を予測する。

ここで、単語を予測する場合、単語間の係り受けは有用な情報である。ところ

が、係り受けの関係のある「何が」は考慮されない。このように、Chelba & Jelinek の方法では、予測に有用な情報を考慮に入れることができないことがしばしば起こる。

そこで本発明は、予測の精度を高めることのできる単語予測方法、装置、さらには音声認識方法、装置等の提供を課題とする。

【0009】

【課題を解決するための手段】

本発明者は、単語予測の各段階でその時点までの単語列を覆う部分解析木の列を履歴情報として利用することに着目した。そして履歴となる部分解析木の列から、次単語の予測により有用な情報をもつ部分解析木を選択する。言い換れば履歴となる単語列および係り受け構造を元に、次単語の予測に最も有用な単語および／または単語列を選択すれば、予測精度を高めることができる。つまり、予測すべき単語と係り受けの関係のある部分解析木を特定し、その後、当該部分解析木、すなわち予測対象の単語と係り受けの関係のある推定された単語および／または単語列から次の単語の予測を行なう。Chelba & Jelinek の方法と異なり、予測される単語を含む文の構造を利用して、予測に有用な情報のみを考慮することができる。

【0010】

本発明は以上の思想に基づくものであり、予測すべき単語を含む複数の単語からなる文の文構造を特定し、特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係を有する単語および／または単語列を選択し、選択された前記単語および／または単語列に基づいて前記予測すべき単語を予測することを特徴とする単語予測方法である。

本発明において、前記単語および／または単語列は、文構造において部分解析木を構成するものであることが望ましい。

また本発明において、選択された前記単語および／または単語列が複数ある場合に、複数の前記単語および／または単語列に基づいて単語の予測を行なうことができる。

【0011】

また本発明は、予測すべき単語と他の単語および／または単語列との係り受け

の関係を特定するステップと、特定された前記他の単語および／または単語列から前記予測すべき単語を予測するステップと、を備えたことを特徴とする単語予測方法を提供する。

本発明の単語予測方法において、前記係り受けの関係は係り受けの方向を含み、前記予測すべき単語がその前に存在する単語に係る場合がある。

また、予測すべき単語と他の単語および／または単語列との係り受けの関係が複数通り存在する場合には、その各々について単語を予測することができる。

【0012】

また本発明では、文頭から j 番目の単語までを含む文の構造を特定するステップと ($j = 0, 1, 2, \dots$) 、前記 j 番目の単語までを含む文の構造から $j + 1$ 番目の単語に係る単数または複数の部分解析木を特定するステップと、前記単数または複数の部分解析木に基づき前記 $j + 1$ 番目の単語を予測するステップと、予測された $j + 1$ 番目までを含む文の構造とその確率値を求めるステップと、以上の各ステップを文末の単語まで実行した後に、前記確率値が最大となる文構造および単語列を音声認識の結果として選択するステップと、を備えたことを特徴とする音声認識方法が提供される。

【0013】

また本発明では、学習されたテキストデータが記述された辞書を記憶した記憶手段と、予測すべき単語を含む複数の単語からなる文の文構造に基づいて前記予測すべき単語と係り受けの関係にある単語および／または単語列を選択し、かつ選択された前記単語および／または単語列に基づいて前記予測すべき単語を予測する予測手段と、を備えたことを特徴とする単語予測装置が提供される。

さらにまた本発明では、学習されたテキストデータが記述された辞書を記憶した記憶手段と、予測すべき単語と係り受けの関係を有する単語および／または単語列を選択し、さらに選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する予測手段と、を備えたことを特徴とする単語予測装置が提供される。

本発明の単語予測装置において、前記辞書は、学習されたテキストデータに基づく所定の単語または単語列に対して他の所定の単語が出現する頻度を記憶する

ことができる。

また、前記予測手段は、前記頻度に基づき予測された単語を含む単語列の確率値を計算することができる。

さらに、前記予測手段は、前記確率値が最大となる単語列を予測結果として選択することができる。

【0014】

また本発明は、アナログ音声入力信号をデジタル信号に変換処理する音響処理手段と、音の特徴を学習した音響モデルを記憶した記憶手段と、所定の単語または単語列に対する他の所定の単語の出現頻度を記述した辞書を記憶した記憶手段と、前記デジタル信号について前記音響モデルおよび前記辞書を用いて確率値を計算して最も確率の高い単語を入力した音声として認識する認識手段と、を備え前記認識手段は、予測すべき単語を当該単語を含む文の構造に基づき予測するとともに、前記出現頻度を用い予測された単語を含む文について前記確率値を計算することを特徴とする音声認識装置を提供する。

さらに本発明は、予測すべき単語と他の単語および／または単語列との係り受けの関係を特定し、前記係り受けの関係を有していると特定された単語および／または単語列を用いて前記予測すべき単語を予測する手段と、前記手段により予測された単語および／または単語列、ならびに係り受けの関係を表示する表示手段と、を備えたことを特徴とするコンピュータ・システムを提供する。

また、本発明は、コンピュータ・プログラムが記憶されかつコンピュータで読み取り可能な記憶媒体であって、前記コンピュータ・プログラムは、予測すべき単語を含む複数の単語からなる文の文構造を特定し、特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係を有する単語および／または単語列を選択し、選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する処理をコンピュータに実行させることを特徴とする記憶媒体を提供する。さらにまた、本発明は、コンピュータ・プログラムが記憶されかつコンピュータで読み取り可能な記憶媒体であって、前記コンピュータ・プログラムは、予測すべき単語と他の単語および／または単語列との係り受けの関係を特定するステップと、特定された前記他の単語および／または単語列か

ら前記予測すべき単語を予測するステップと、をコンピュータに実行させるものであることを特徴とする記憶媒体を提供する。

また本発明は、コンピュータに、予測すべき単語を含む複数の単語からなる文の文構造を特定し、特定された前記文構造に基づき、前記予測すべき単語と係り受けの関係を有する単語および／または単語列を選択し、選択された前記単語および／または前記単語列に基づいて前記予測すべき単語を予測する処理を実行するためのプログラムを記憶する記憶手段と、前記記憶手段からプログラムを読み出して前記プログラムを送信する送信手段と、を備えたことを特徴とするプログラム伝送装置を提供する。さらにまた、本発明は、コンピュータに、予測すべき単語と他の単語および／または単語列との係り受けの関係を特定するステップと、特定された前記他の単語および／または単語列から前記予測すべき単語を予測するステップとを実行するためのプログラムを記憶する記憶手段と、前記記憶手段からプログラムを読み出して前記プログラムを送信する送信手段と、を備えたことを特徴とするプログラム伝送装置を提供する。

【0015】

【発明の実施の形態】

以下本発明の実施の形態を図面を参照しつつ説明する。

図1は、本実施の形態による音声認識装置を含む構成を示すブロック図である。

ブロック101において生成された文章（真の文章）としての単語列Wは、Sとして発話される（ブロック102）。発話されたSは音声認識手段110を構成する音響処理部111に入力される。音響処理部111では入力された発話Sを信号Xに変換するとともに、記憶する。変換された信号Xは、言語復号部112において認識結果W'となる。言語復号部112は、音の特徴を学習した音響モデル113と、やはり学習により予め作成されたテキストデータを有する辞書114とを備えている。認識された結果W'の文章は表示される（ブロック120）。

【0016】

図2は本実施の形態による音声認識装置を実現するためのシステムの一例を示

している。このシステムは、マイク210、コンピュータ装置220および表示装置230から構成される。コンピュータ装置220は、サウンドカード221、メモリ222およびCPU223を備えている。

図2に示すシステムにおいて、発話者が発話した音声はマイク210にアナログ信号として入力される。このアナログ信号は、コンピュータ装置220のサウンドカード221によりデジタル信号に変換処理された後にメモリ222に記憶される。このメモリ222には、前述の音響モデル113および言語モデルを含む辞書114も記憶されている。CPU223は、メモリ222に記憶されているデジタル信号および辞書に基づき言語の復号を行なう。また、以下説明する単語予測方法を実現するためのプログラムを解釈実行する。復号された言語が認識結果であり、この認識結果は、表示装置230に表示される。このプログラムは、メモリ222に記憶されている。

以上のシステムにおいてマイク210はコンピュータ装置220と別体として示しているが、コンピュータ装置220あるいは表示装置230と一体化していても良い。つまり発話された音声を入力するためのマイクが存在すれば、その形態は問わない。また、認識結果を表示装置230、例えばCRTに対して行なう例を示しているが、プリンタにより印刷物として提供する、あるいはフレキシブル・ディスクその他の記憶媒体に記憶させることもできる。

本実施の形態では、以下説明する単語予測方法の前提として、音響処理部111で変換された音声信号について音響モデル113を用いた計算結果として認識候補として単語を選択する。この選択された単語について、以下説明する単語予測がなされ、最終的に音声認識の結果が得られる。

【0017】

本実施の形態に用いる単語予測方法を説明する。

本発明は、前述の様に、予測すべき単語と係り受けの関係のある部分解析木を予測し、その後、当該部分解析木、すなわち次の単語に係ると推定された先行単語および／または単語列から次の単語の予測を行なうことを提案する。換言すれば、予測すべき単語と係り受けの関係のある部分解析木から次の単語を予測する。

tri-gramおよびChelba & Jelinek の方法を説明するために用いた、「さらに小選挙区の」という例文に基づき、本実施の形態を説明する。

「さらに小選挙区の」という文章は、「さらに」、「小」、「選挙」、「区」、および「の」という5つの単語から構成される。ここで、文頭から数えて何番目の単語であるかをjで表すことになると、「の」は、5番目の単語である。また、5番目の単語である「の」までを含む文の構造は、図3 (a) ~ (c) に示すように3通り存在するものとする。なお、文の構造とは、各単語間の係り受けの関係をいうものとする。この3通りについて説明すると以下の通りである。

【0018】

図3 (a) は、「さらに」は「の」に係らないが、「区」は「の」に係るケースを示している。この様子を図3 (a) では矢印で示しており、「さらに」から引き出されている矢印は「の」以降の単語を指し、また「区」から引き出されている矢印は「の」を指している。ここで、部分解析木としてみると、「さらに」が1つの部分解析木を、また、「小選挙区の」がもう1つの部分解析木を構成するから、図3 (a) の場合には、「の」に係る部分解析木は、「小選挙区の」のみである。なお、この構造の確率値を0. 034とする。

図3 (b) は、「さらに」および「区」の両者が「の」に係らないケースである。したがって、「さらに」および「区」から引き出されている矢印は「の」以降の単語を指している。この文構造の確率値を0. 001とする。

また、図3 (c) は、図3 (a)、(b) の「の」が類似音の「を」と予測した例である。この「を」の予測は音響モデル113により指示される。文構造としては図3 (a) と同様に、「さらに」は「の」に係らないが、「区」は「の」に係るものとしている。図3 (c) のケースの確率値を0. 028とする。つまり、図3 (a) の確率値が最も高い。この段階では、確率値の最も高い図3 (a) のケースが音声認識結果の候補ということができる。

なお、以上の図3 (a) ~ (c) は、本実施の形態を説明するための限られた例であり、例えば、5番目の単語が「を」の場合に、係り受けとしては図3 (b) と同様なケースも存在し得るし、また、5番目の単語が「の」、「を」ではなく「と」であるケースも存在し得る。いずれにしても、図3 (a) ~ (c) は、

j番目（ここでは、5番目）までの単語を含む構造と確率値を求めた結果を示している。

なお、図3中の、s [5] [0] の、[5]は5番目の単語が処理の対象となっていることを示している。また、[0]は3通りある係り受けの関係のうちのいずれが処理の対象となっているかを示している。

【0019】

次に、6番目の単語を予測する。この6番目の単語を予測するに当たって、6番目の単語を含む文構造をまず特定する。図3(a)の例では、「の」のみが6番目の単語に係るケース、「さらに」および「の」の両者が6番目の単語に係るケース、「さらに」および「の」の両者が6番目の単語に係らないケース、の3通りが存在する。この3通りを特定し、各々について、6番目の単語の予測を行なう。それぞれのケースを図3(a-1)～(a-3)に示すが、本実施の形態では、予測対象となる6番目の単語を予測する前に、6番目の単語を含めた文構造が特定されることになる。

ここで、辞書114には、学習されたテキストデータに基づく、所定の単語または単語列に対する他の単語の出現頻度が記述されている。例えば、テキストデータ中に、「小選挙区の」という文章がn回出現し、かつその後に「導入」という単語が出現した回数がm回あったとすると、「小選挙区の」に対する「導入」の出現頻度は m/n ということになる。また、「さらに」および「小選挙区の」の2つの部分解析木から「導入」を予測する場合には、「さらに」と「小選挙区の」が出現した後に「導入」が出現する頻度を考慮する必要がある。つまり、テキストデータ中に、「さらに」と「小選挙区の」が出現した文章がn'回出現し、かつその後に「導入」という単語が出現した回数がm'回あったとすると、「さらに」と「小選挙区の」に対する「導入」の出現確率は m'/n' ということになる。このとき、経験的に「さらに」は、動詞や形容詞などの用言に高い頻度で係り、名詞などの体言に係ることは稀なので、名詞である「導入」が出現する回数m'は非常に小さいであろうから、「さらに」および「小選挙区の」の2つの部分解析木から「導入」を予測した際の確率値は、「小選挙区の」のみから予測する場合と比較して相当低くなる。換言すれば、「導入」を予測する場合には、

「さらに」を考慮するのは望ましくないということである。

したがって、「の」から「導入」を予測したときの「さらに小選挙区の導入」という文の確率値は、「さらに」と「の」とから「導入」を予測したときの「さらに小選挙区の導入」とい文の確率値より高くなる。

なお、ここでは、図3 (a-1)、(a-2)について説明したが、図3 (a-3)のケースについても同様に確率値が計算される。また、文末の単語まで同様に予測がなされることになる。

【0020】

次に、図4を参照しつつ、図3 (a-1)のケースの単語予測過程を説明する。

図4 (a) は、図3 (a-1) の状態を示している。この状態から、「?」で示す次の単語（この例では6番目の単語）と係り受けの関係にある部分解析木を特定する。このケースでは、「小選挙区の」からなる部分解析木は6番目の単語に係るが、「さらに」からなる部分解析木は6番目の単語に係らない。この係り受けの様子を図4 (b) に示している。つまり、「さらに」から引き出される矢印は6番目以降の単語を指しており、6番目の単語と「さらに」が係り受けの関係にないことを示している。また、「小選挙区の」中の「の」から引き出される矢印は6番目の単語を示す「?」を指しており、6番目の単語に「小選挙区の」という単語列が係ることを示している。

以上のように、予測すべき6番目の単語を含めた文構造を予測した後に、6番目の単語と係り受けの関係にある部分解析木「小選挙区の」から「導入」を予測する。また、「導入」を予測後に、図4 (c) に示すように、「導入」を含めた文構造の予測を行なう。つまり、図3 (a-1) のケースによれば、予測精度を下げる可能性の高い「さらに」を考慮しないので、高い確率値を得ることができる。

【0021】

以上本実施の形態による単語予測方法を説明したが、次に図5に示すフローチャートを参照しつつ最終的に音声認識結果を出力するまでの流れを説明する。なお先にも述べたが、音響処理部111で変換された音声信号について音響モデル

113を用いた計算結果として認識候補として単語を選択し、この選択した単語についてさらに予測による絞込みを行なうのが、以下の処理である。

図5において、はじめに、何番目の単語について処理するか(S100)、また、何番目の構造を処理するか(S101)を確定する。何番目の単語を処理するかはjで示され、また、何番目の構造を処理するかはiで示される。文頭から予測を行なうため、j、iの初期値は0である。なお、j、iの具体的な態様は、図3を参照することにより容易に理解できる。

次いで、予測すべき単語を含む文の構造と確率値を求める(S102)。ここで、図5中、S104のs[j][]は、j番目の単語までを含む文構造とその確率値を示すものである。図3の例では、はじめに、s[5][0]、つまり5番目の単語について、3通りのうちの1番目の文構造とその確率値を求めている。この文構造および確率値は、次の単語の予測に用いられるから、s[j+1][]に対して列挙される(S102)。図3の例では、はじめに図3(a-1)がs[6][]に対して列挙されている。

文構造が複数通り存在する場合に、そのすべての構造についてS102の処理を行なう。i = i + 1を内容とするS103、および、すべてのs[j][]をチェックしているか否かを判定するS104により、これを実現する。

すべての構造についてS102の処理が完了したならば、次の単語について同様の処理を実行する。次の単語は、j = j + 1で定義される(S105)。j = j + 1が当該文書の最後の単語でなければ、S101からの一連の処理を実行する。また、最後の単語であれば、s[j][]から確率値が最大の文構造と単語列を選択し、表示装置230に表示する。この文構造は、係り受けの関係を矢印で示したものとすることができますし、部分解析木の構造とすることもできる。

以上の実施形態では、パーソナルコンピュータにおいて本発明を実現した例を示した。しかし、本発明は、所定のプログラムを格納した記憶媒体、あるいはプログラムを伝送するための伝送装置として実現することもできる。

【0022】

次に、「何がいま政治の再生のために」という例文に基づいて本発明を説明する。

「何がいま政治の再生のために」という文章は、「何」、「が」、「いま」、「政治」、「の」、「再生」、「の」、「ため」および「に」という9つの単語から構成されている。また、「何が」、「いま」および「政治のために」という3つの部分解析木から構成されている。

図6 (a) は「何がいま政治の再生のために」までの単語の予測が完了した状態を示している。前述のように、本文章は、「何が」、「いま」および「政治の再生のために」という3つの部分解析木から構成されている。そして、「何が」という部分解析木は、「何」が「が」に係ることまで予測しているが、「が」、換言すれば「何が」からなる部分解析木の係る単語については未予測の状態である。この状態は、図6 (a) の「が」から引き出されている矢印が「?」を示していることで理解できる。また、「いま」からなる部分解析木、さらに「政治の再生のために」からなる部分解析木の係る単語についても未予測の状態である。

【0023】

図6 (a) の状態から、「?」で示す次の単語（この例では10番目の単語）と係り受けの関係にある部分解析木を予測する。この例文では、「何が」、「いま」および「政治の再生のために」という3つの部分解析木が全て10番目の単語に係ることを予測、特定している。この係り受けの関係を図6 (b) に示している。つまり、「何が」の「が」、「いま」および「政治の再生のために」の「に」から引き出された矢印は、10番目の単語を指している。

以上のように、予測すべき10番目の単語を含めた文構造を特定した後に、10番目の単語を予測する。つまり、3つの部分解析木（「何が」と「いま」と「政治の再生のために」）が全て予測すべき単語に係るので、これら全てを考慮して「必要」を予測するのである。

Chelba & Jelinekの方法では、「ため」と「に」とから「必要」を予測しているが、「何が」は「必要」を予測する上で有用な情報であるから、その点で本実施の形態の予測精度が高いということができる。

【0024】

以上では日本語の例文について説明したが、次に英語の例文について説明する。

日本語と英語との構文上の相違の一つとして、日本語では係り受けの方向が一定であるのに対して英語では係り受けの方向が一定でないことがある。本実施の形態を英語などの係り受けの方向が一定でない言語に適用する場合には、次の単語と係り受けの関係にある部分解析木とその係り受けの方向を特定し、その後、係り受け関係にある部分解析木から次の単語の予測を行なえばよい。

例は、「the contact ended with a loss」から「after」を予測するものである。「the contact ended with a loss」という文章は、「the」、「contact」、「ended」、「with」、「a」および「loss」という6つの単語から構成される。そして、部分解析木としてみると、「the contact」が1つの部分解析木を、また、「ended with a loss」が1つの部分解析木を構成する。

図7(a)は、「the contact ended with a loss」までの単語の予測が完了した状態を示している。前述のように本文章は、「the contact」および「ended with a loss」という2つの部分解析木から構成されている。そして、図7(a)の矢印が示すように、「the contact」という部分解析木においては「the」が「contact」に係る。また、「ended with a loss」という部分解析木においては、「a」が「loss」に係り、「loss」が「with」に係り、さらに「with」が「ended」に係る。このように、英語は、前方から後方へ、後方から前方へと二方向の係り受けの関係を有している。

【0025】

図7(a)の状態から、「?」で示す次の単語（この例では7番目の単語）に係る部分解析木を予測する。つまり、図7(b)に示すように、7番目の単語が、「ended」に係ることを予測する。「ended」は「ended with a loss」という部分解析木に含まれるから、7番目の単語は「ended with a loss」に係ることに基づき予測されたことになる。次に、図7(c)に示すように、この「ended with a loss」から構成される部分解析木から、「after」を予測する。

【0026】

＜実験例＞

約1000文のモデルを新聞記事に基づき作成した。このモデルに本実施の形態の手法を用いてクロスエントロピーを求める実験を行なった。その結果は以下

の通りであった。

本実施の形態 4.05 [bit]

tri-gram 4.27 [bit]

ここで、本実施の形態の4.05 [bit] という値は16.6個の単語からの選択に相当し、また、4.27 [bit] という値は19.3個の単語からの選択に相当する。したがって、本実施の形態により、単語予測精度の向上がなされたことが確認できた。

【0027】

【発明の効果】

以上説明したように、本発明によれば、予測すべき単語を含む文構造を特定し、かつ予測すべき単語と係り受けのある単語または単語列から単語予測を行なう。係り受けの関係は、単語予測に有効な情報であるから、単語予測の精度が高くなる。

【図面の簡単な説明】

【図1】 本実施の形態を示すブロック図である。

【図2】 本実施の形態によるコンピュータ・システム構成を示す図である

【図3】 本実施の形態による単語予測を説明するための図である。

【図4】 本実施の形態による単語予測を説明するための図である。

【図5】 本実施の形態による音声認識を説明するためのフローチャートである。

【図6】 本実施の形態による単語予測を説明する他の例を示す図である。

【図7】 本実施の形態による単語予測を説明する他の例を示す図である。

【図8】 tri-gramモデルによる単語予測を説明するための一例を示す図である。

【図9】 tri-gramモデルによる単語予測を説明するための他の例を示す図である。

【図10】 Chelba & Jelinekの方法による単語予測を説明するための一例を示す図である。

【図11】 Chelba & Jelinekの方法による単語予測を説明するための他の

例を示す図である。

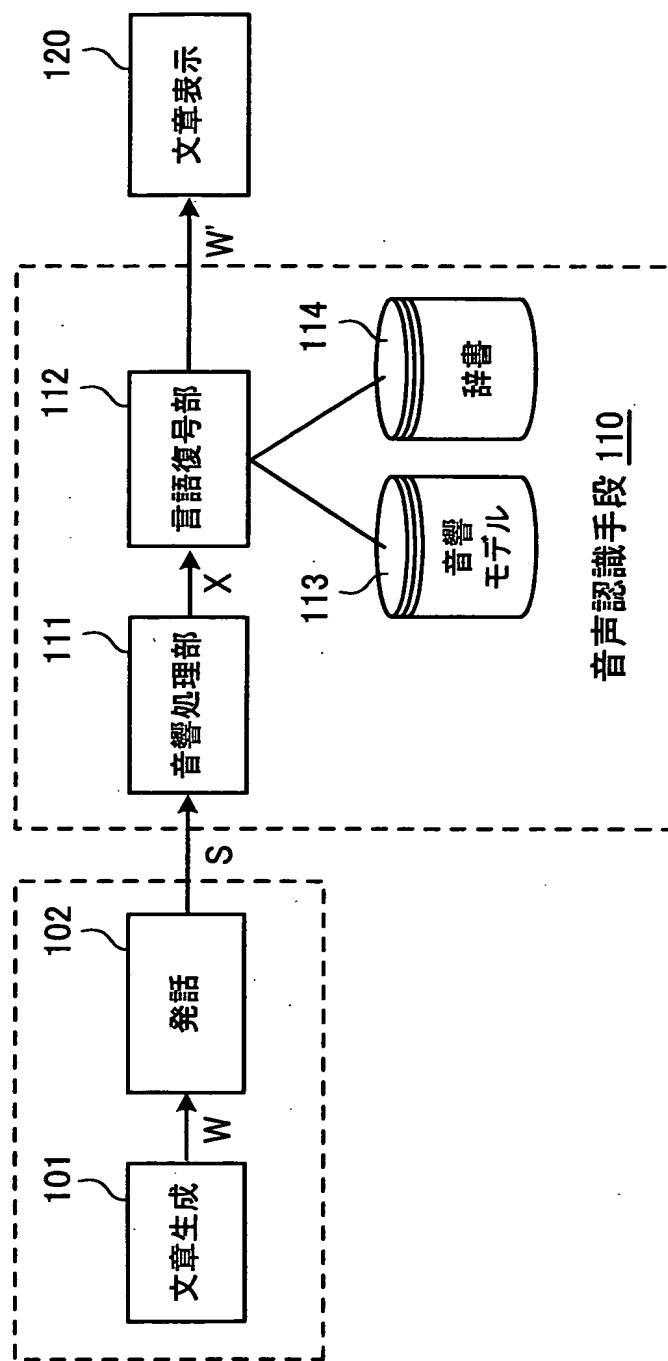
【符号の説明】

110…音声認識手段、210…マイク、220…コンピュータ装置、221…
サウンドカード、222…メモリ、223…C P U、230…表示装置

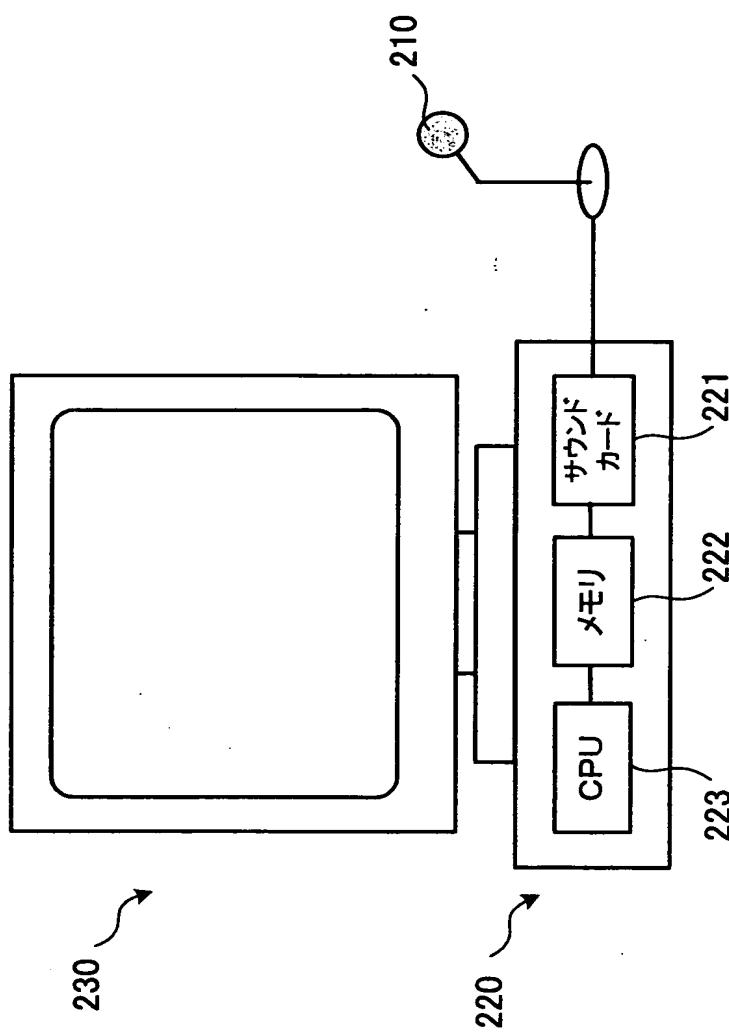
【書類名】

図面

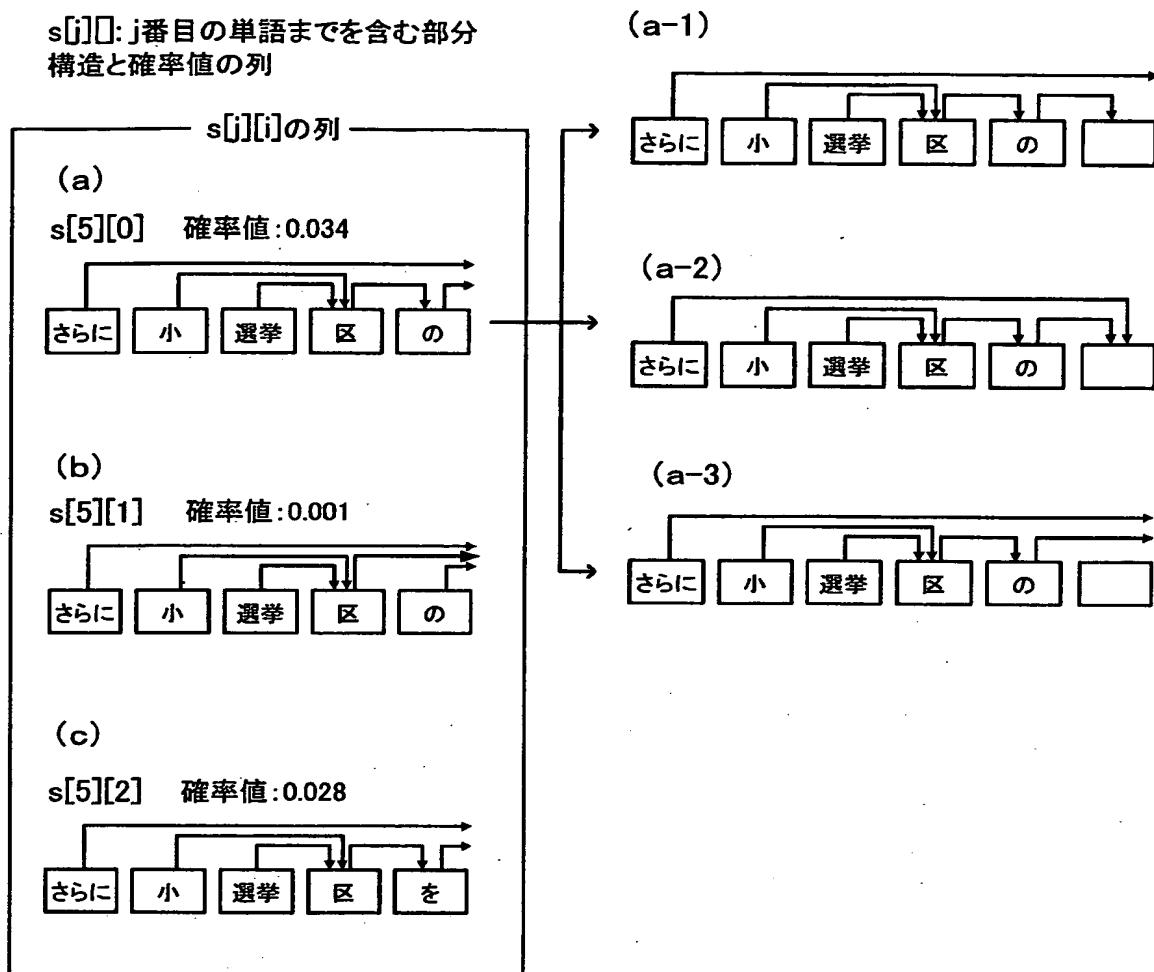
【図1】



【図2】

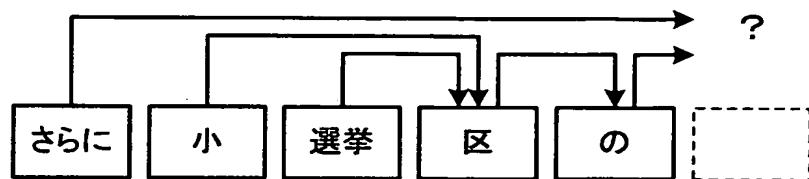


【図3】

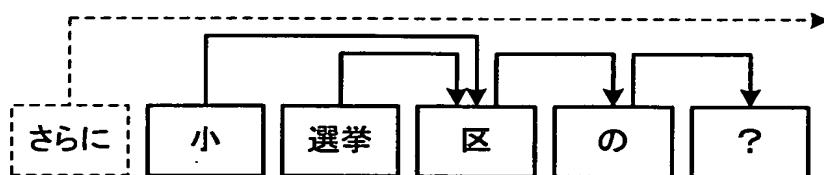


【図4】

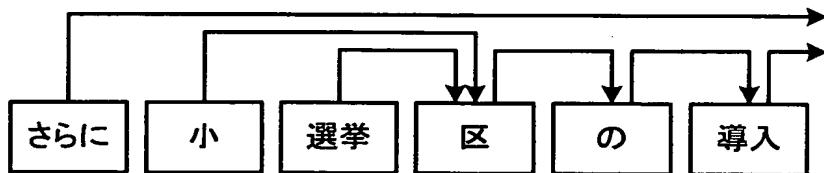
(a)



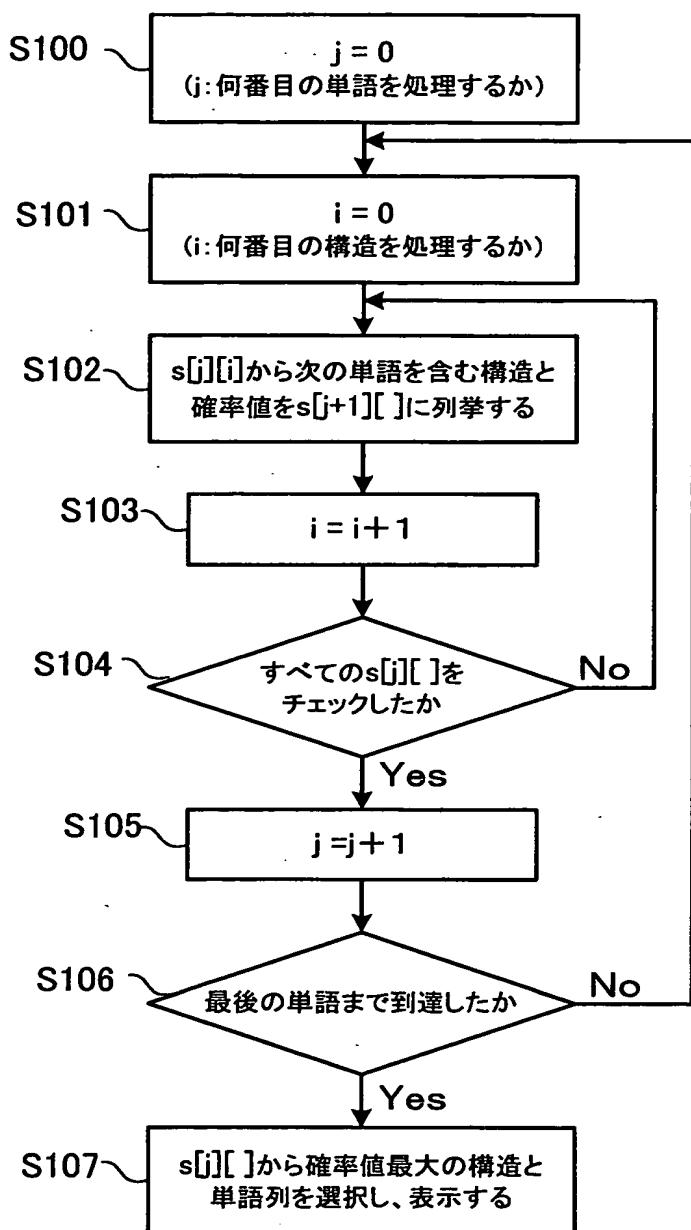
(b)次の単語に係る部分解析木を予測



(c)係る部分解析木から次の単語を予測

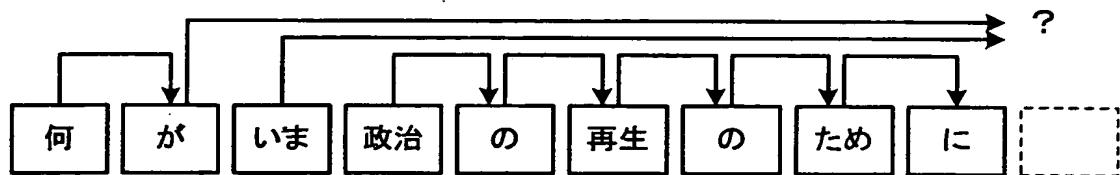


【図5】

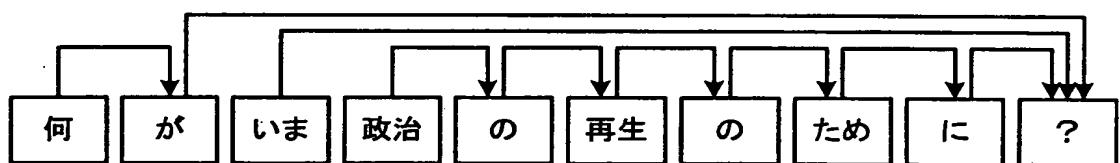


【図6】

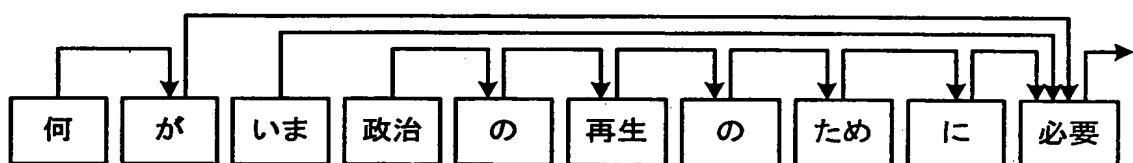
(a)



(b)次の単語に係る部分解析木を予測

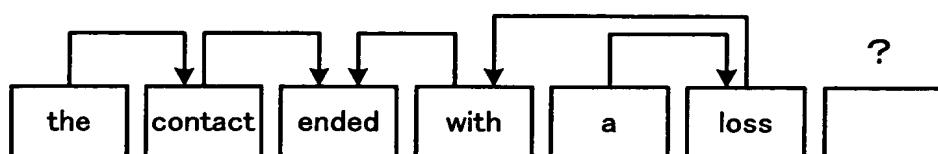


(c)係る部分解析木から次の単語を予測

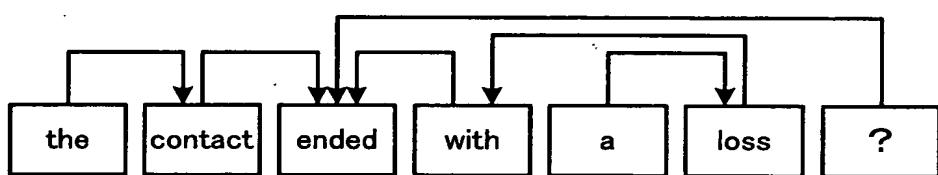


【図7】

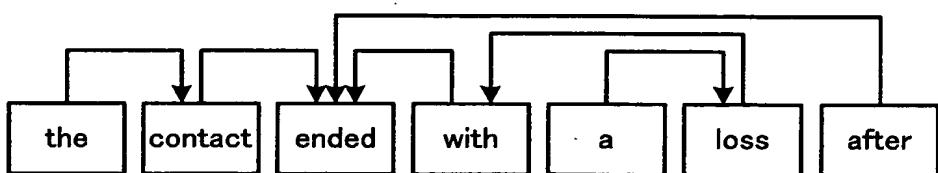
(a)



(b)次の単語に係る部分解析木を予測



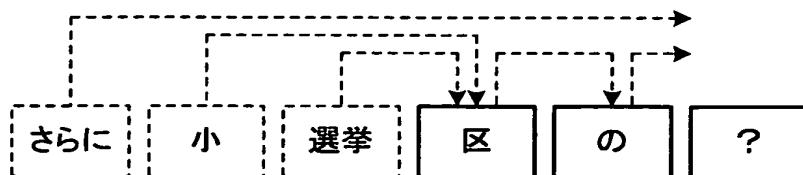
(c)係る部分解析木から次の単語を予測



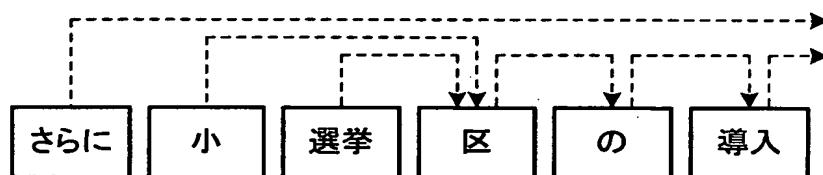
【図8】

tri-gramモデル

(a)



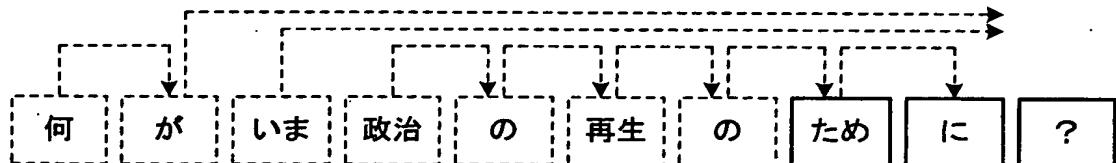
(b)直前の2単語から次の単語を予測



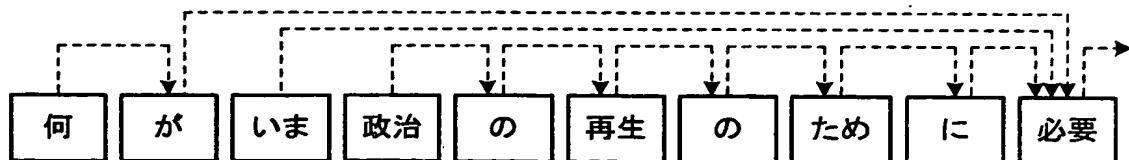
【図9】

tri-gramモデル

(a)



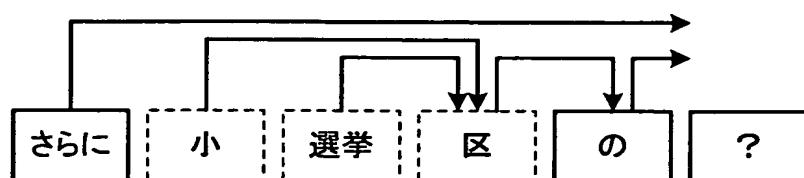
(b)直前の2単語から次の単語を予測



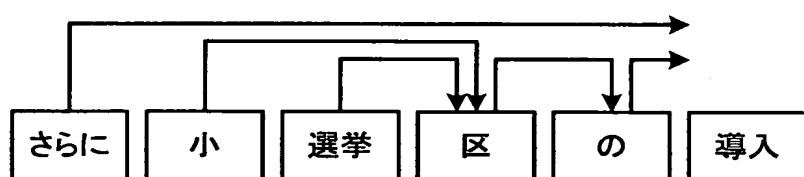
【図10】

Chelba & Jelinekの方法

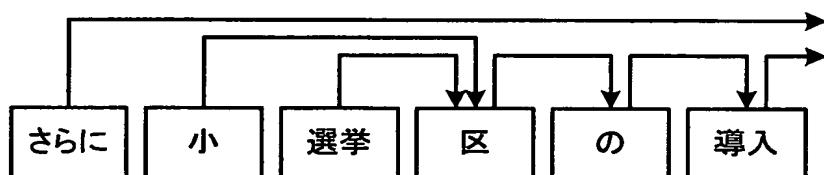
(a)



(b)直前の2つの主辞から次の単語を予測



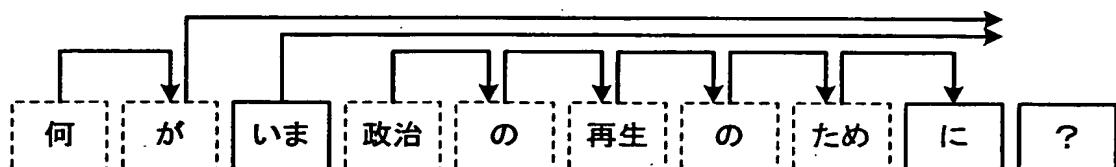
(c)次の構造を予測



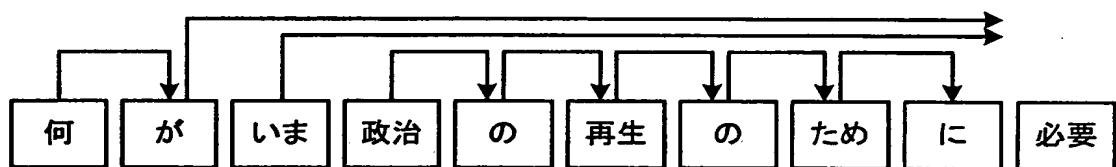
【図11】

Chelba & Jelinekの方法

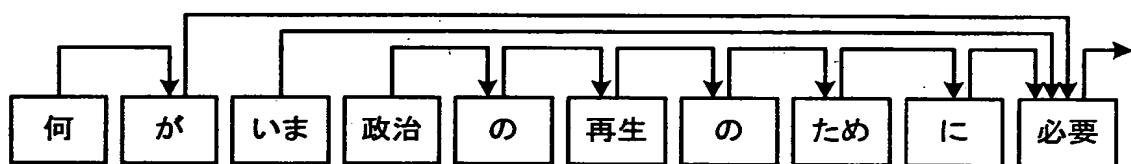
(a)



(b)直前の2つの主辞から次の単語を予測



(c)次の構造を予測



【書類名】 要約書

【要約】

【課題】 予測の精度を高めることのできる単語予測方法を提供する。

【解決手段】 「？」で示す6番目の単語を予測する場合に、6番目の単語と係り受けの関係を有する部分解析木を予測する。「さらに小選挙区の」は、「さらに」および「小選挙区の」という2つの部分解析木を有している。「さらに」は6番目の単語と係り受けの関係を有さず、一方、「小選挙区の」は6番目の単語と係り受けの関係を有していると予測する。そして、「小選挙区の」から6番目の単語である「導入」を予測する。この例では、「さらに」は「導入」の予測に有効な情報とならないため、「小選挙区の」のみから予測することが望ましい。

【選択図】 図4

認定・付加情報

特許出願の番号 特願2000-210599
 受付番号 50000874473
 書類名 特許願
 担当官 宇留間 久雄 7277
 作成日 平成12年 8月22日

<認定情報・付加情報>

【特許出願人】

【識別番号】 390009531
 【住所又は居所】 アメリカ合衆国 10504、ニューヨーク州 アーモンク (番地なし)
 【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】 100086243
 【住所又は居所】 神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内

【氏名又は名称】 坂口 博

【復代理人】

【識別番号】 100104880
 【住所又は居所】 東京都港区赤坂 5-4-11 山口建設第2ビル
 6F セリオ国際特許事務所
 【氏名又は名称】 古部 次郎

【選任した代理人】

【識別番号】 100091568
 【住所又は居所】 神奈川県大和市下鶴間 1623 番地 14 日本アイ・ビー・エム株式会社 大和事業所内
 【氏名又は名称】 市位 嘉宏

【選任した代理人】

【識別番号】 100106699
 【住所又は居所】 神奈川県大和市下鶴間 1623 番 14 日本アイ・ビー・エム株式会社 大和事業所内
 【氏名又は名称】 渡部 弘道

【選任した復代理人】

【識別番号】 100100077

次頁有

認定・付加情報（続き）

【住所又は居所】 東京都港区赤坂5-4-11 山口建設第2ビル
6F セリオ国際特許事務所
【氏名又は名称】 大場 充

次頁無

出願人履歴情報

識別番号 [390009531]

1. 変更年月日 2000年 5月16日

[変更理由] 名称変更

住 所 アメリカ合衆国 10504、ニューヨーク州 アーモンク (番地なし)

氏 名 インターナショナル・ビジネス・マシンズ・コーポレーション